(51) International Patent Classification 6 :	A1	(11) International Publication Number:	WO 97/18326
C12Q 1/68		(43) International Publication Date:	22 May 1997 (22.05.97)

(21) International Application Number: PCT/IB96/01219

(22) International Filing Date: 13 November 1996 (13.11.96)

(30) Priority Data:
60/006,632 13 November 1995 (13.11.95) US
60/006,678 14 November 1995 (14.11.95) US

(71) Applicant (for all designated States except US): INSTITUT PASTEUR [FR/FR]; 28, rue du Docteur-Roux, F-75015 Paris (FR).

(72) Inventors; and
(75) Inventors/Applicants (for US only): BENSIMON, Aaron [IL/FR]; 5, rue du Clos-de-l'Abbaye, F-92160 Antony (FR). CREMER, Thomas [DE/DE]; University of Heidelberg, Laboratorium Für Molekulare Cytogenetik, Institut für Humangenetik und Anthropologie, Im Neuenheimer Feld 328, D-69120 Heidelberg (DE). KRAUS, Jürgen [DE/DE]; University of Heidelberg, Laboratorium Für Molekulare Cytogenetik, Institut für Humangenetik und Anthropologie, Im Neuenheimer Feld 328, D-69120 Heidelberg (DE). LICHTER, Peter [DE/DE]; Organisation Komplexer Gerome, Deutsches Krebsforschungszentrum (DKFZ), Im Neuenheimer Feld 280, D-69120 Heidelberg (DE).

(74) Agents: MARTIN, Jean-Jacques et al.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).

(81) Designated States: AU, BR, CA, JP, KR, MX, NZ, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: ULTRAHIGH RESOLUTION COMPARATIVE NUCLEIC ACID HYBRIDIZATION TO COMBED DNA FIBERS

(57) Abstract

A method for detecting and quantifying the presence of a nucleic acids sequence in a genome comprising: (A) immobilizing by combing the target nucleic acids sequence on a supportive matrix, (B) contacting the combed nucleic acids sequence with a 1:1 mixture of differently labeled nucleic acids probe to be tested and reference nucleic acids probe to form hybridization, (C) counting individual labeled hybridization dots located on combed nucleic acids sequence, (D) determining the number of the nucleic acids sequence to be detected and quantified by the calculation of the ratio between the total dot number from the test probe and the total dot number from the reference probe.

ULTRAHIGH RESOLUTION COMPARATIVE NUCLEIC
ACID HYBRIDIZATION TO COMBED DNA FIBERS

BACKGROUND OF THE INVENTION

5   This invention relates to the detection and quantification
of the presence of a gene in a genome. In one embodiment, the
invention relates to the detection and quantification of a
human oncogene.

A recently developed method, called comparative genomic
10   hybridization (CGH) (Kallioniemi et al. 1992, du Manoir et al.
1993; Joos et al. 1993) has provided a new tool to detect non-
random gains and losses of DNA sequences in genomic DNA
(obtained e.g., from tumor specimens). For CGH, genomic tumor
DNA is labeled with a hapten (e.g., Biotin) or directly with a
15   fluorochrome (e.g., FITC). Genomic DNA prepared from normal
cells (of the patient or other persons) is differently labeled
with another hapten (e.g., digoxigenin) or directly with
another fluorochrome (e.g., TRITC or Texas red). Labeled tumor
and control DNAs are mixed in equal amounts. This mixture is
20   hybridized in the presence of an excess of unlabeled cot1-DNA
to normal metaphase spreads (target chromosomes) prepared from
a healthy male or female person. The excess of cot1-DNA
hybridizes to labeled interspersed and tandem repetitive
sequences present in genomic DNA as well as to such sequences
25   present in the target chromosomes. This step is essential to
suppress the unwanted hybridization of the repetitive sequences
to the target chromosomes (Lichter et al. 1988, Pinkel et al.
1988). Following hybridization and washing steps, standard
detection procedures are applied to visualize haptenized
30   sequences with two different fluorochromes (Lichter and Cremer
1992). This step is omitted in case of DNA-probes directly
labeled with fluorochromes.

Consider for example a tumor with an essentially diploid
karyotype except for a few monosomic or trisomic chromosomes or
35   chromosome segments. Labeled DNA fragments with a size of

several hundred base pairs from the tumor DNA and the normal control DNA will hybridize with equal probability to their respective target sequences. The labeled DNA fragments from a chromosome or chromosome segment present in two copies in the pseudodiploid tumor cells together with the differently labeled fragments from genomic DNA of diploid cells yield a certain color mixture on the respective target chromosome or chromosome segment, e.g., yellow when the chromosome or chromosome segment is labeled with equal numbers of green and red fluorochromes representing the hybridized fragments from tumor DNA and normal DNA, respectively. For a chromosome segment present in three or higher copy numbers in the tumor, this color would become more greenish, while the loss of the segment would result in a more red color.

These color changes can be quantitatively recorded by measuring fluorescence ratio profiles along target chromosomes (du Manoir 1994, Piper et al. 1994). The choice of the appropriate equipment to measure signal intensities is important. Detectors should allow linear intensity measurements over a wide range. CCD-camerase are particularly useful in this respect. All data are stored digitally so that they can be used by microprocessor for the calculation of fluorescence ratios. In this way, a copy number karyotype can be established.

CGH can be performed with DNA extracted from archived, paraffine embedded tissues. Even minute amounts of genomic DNA can be used for this purpose after amplification with degenerate oligonucleotide primers (DOP-PCR) (Telenius et al. 1992; Speicher et al. 1993; Isola et al. 1994). It is possible to microdissect areas containing tumor cells from a tissue section and screen it for gains and losses of genetic materials by CGH performed with DOP-PCR amplified DNA (Speicher et al. 1994).

The number of studies, which demonstrate the usefulness of CGH to detect DNA copy number changes in tumors, is rapidly

increasing. Studies published so far reflect already a variety of tumors, including various forms of acute and chronic leukemias, bladder cancer, breast cancer, colorectal cancer, gliomablastoma, kidney cancer, neuroblastomas, prostate cancer, small cell and non-small cell lung carcinomas, uvea melanomas (e.g., du Manoir et al. 1993, 1994; Isola et al. 1994a,b; Joos et al. 1993; Kallioniemi et al. 1992, 1993, 1994a,b; Muleris et al. 1994; Ried et al. 1994; Schrock et al. 1994, Speicher et al. 1993, 1994, 1995 and our unpublished data). Numerous hitherto unknown regions, in particular amplification sites, have been found in these studies and will become the focus of efforts to clone the respective tumor relevant genes. Different tumor entities generally show distinctly different patterns of non-random changes and clinical follow up studies will show to which extent specific gains and losses can be correlated with the clinical course and prognosis of a given tumor.

The minimum size of a chromosome segment for which a single copy number change can be detected at present by CGH is in the order of 10 Mbp (Joos et al. 1993; du Manoir et al. 1994; Piper et al. 1994). Possibly, the resolution can be somewhat improved, when CGH is performed on prometaphase chromosomes. For amplified DNA sequences the detection limit of CGH is presently about 2 Mbp (number of amplification repeats times amplicon size). Still the precision with which the borders of chromosome segments involved in gains or losses is limited by the banding resolution of the target chromosomes. Thus, the current status of the CGH development does not allow to define the copy number representation of single tumor relevant genes (e.g., oncogenes, tumor suppressor genes). This limits the application of CGH to the detection of gross, unbalanced chromosomal abnormalities.

SUMMARY OF THE INVENTION

This invention aids in fulfilling these needs in the art by providing a method of ultrahigh comparative genomic

hybridization, which differs from state-of-the-art comparative
genomic hybridization by the following new and essential
features. The approach is performed on combed DNA fibers
instead of reference chromosomes and referred to as combed
fiber CGH. Comber fiber CGH allows the analysis of copy number
representation of specific sequences (represented by the combed
DNA fibers) in a genomic test DNA with an ultrahigh resolution
(in the kb-pair range instead of the Mb-pair range as
previously published methods). This improvement makes combed
fiber CGH a very useful method to study the copy number
representation of single genes or parts thereof.

Combed fiber CGH is particularly suited to eliminate
background problems in fluorescence measurements, which arise
when the fluorescence is measured from entire DNA-spots. In
the case of combed fiber CGH, the area of fluorescence
measurements is adapted to a single fiber in a way that only
hybridization dots located precisely on the DNA fiber
contribute to the measured signal derived from the test or
reference genomic DNAs. This improvement provides a very
considerable advantage of combed fiber CGH as compared to a CGH
approach where fluorescence measurements are obtained from
entire DNA-spots of non-ordered target DNA sequences attached
to a supportive matrix.

In addition to fluorescence intensity measurements, which
can be carried out on individual combed target DNA fibers by
standard procedures, combed fiber CGH allows the counting of
hybridization dots located on the combed DNA fibers. By this
approach the total number of dots on a sufficiently large
series of target DNA fibers resulting from hybridized test and
a reference genomic DNA can be counted and a ratio (or a
difference) between the total dot number from the test genomic
DNA and the total dot number from the reference genomic DNA can
be calculated as a measure of the copy number representation of
the combed target DNA sequence in the test and reference
genomic DNA. This approach has not be realized so far in CGH

experiments and allows the evaluation of combed fiber CGH
experiments in situations where the resulting signals are too
weak to allow meaningful measurements of fluorescence ratios on
single DNA fibers.

5   Instead of 1:1 mixtures of test and reference genomic
DNAs, 1:1 mixtures of RNA preparations (or from cDNAs
synthesized from RNA preparations) representing equivalent
numbers of test and reference cells can be hybridized under
suppression conditions (see standard CGH procedures) to combed
10   target DNA fibers e.g., cDNA fibers representing the coding
sequence of genes of interest. The evaluation of the
experiment is performed as described above. This approach
allows an estimate of the relative copy number representation
of mRNAs in test cells (e.g., tumor cells) and reference cells
15   (e.g., normal progenitor cells of the tumor cells).

Target DNA fibers, e.g., cosmids containing (part of) a
gene of interest are genetically engineered in a way that
interspersed repetitive sequences are removed to avoid problems
of insufficient suppression hybridization. (Note: only target
20   fiber specific single copy sequences, but not interspersed
repetitive sequences contained in the hybridization mixture,
can hybridize to the combed target DNA fibers under these
precautions.)

In case the localization of individual target DNA fibers
25   cannot be easily identified by arrays of numerous signal dots
along each fiber, several procedures can be followed for an
unequivocal target fiber identification with fluorochromes that
have an emission spectrum, which allows a clear distinction
from the emission spectrum of the fluorochromes implied in the
30   visualization of hybridized test and reference genomic DNA
fragments. a) Fibers can be stained with appropriate
fluorescent DNA stains. b) Target fiber DNA can be cloned in
the presence of fluorochrome labeled DNA nucleotides for the
direct visualization of the fiber or in the presence of hapten
35   modified nucleotides, e.g., BRdU, biotin, digoxigenin, for

indirect visualization by, e.g., indirect immunofluorescence.
c) Target fiber DNA can be visualized by the addition of an
appropriate amount of labeled target DNA sequences to the
hybridization mixture employed for combed fiber CGH. In this
case, the amount of labeled target sequences should not be
excessive in a way that suppresses the hybridization of target
specific labeled test and genomic DNA sequences. d) Target
fiber DNA containing vector sequences can be hybridized with
labeled vector sequences added to the hybridization mixture.
If desirable, linker DNA of different length can be added to
target fiber DNA in a way that makes it possible to identify
the course and orientation of the combed target fiber by
signals derived from the addition of labeled linker DNA
sequences to the hybridization mixture.

Instead of fluorochromes with different emission spectra,
fluorochromes that differ in fluorescence lifetime can be used.
The choice of fluorochromes is performed with particular
reference to avoid background fluorescence from the supportive
matrix as much as possible.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENT

This invention involves the use of a DNA alignment method
for the detection and quantification of multiple copies of a
gene present in a genome. By multiple copies is meant
visualization of at least about 100 copies per genome, with
very good results at about 2000 copies per genome.

Ultrahigh Resolution Comparative

Nucleic Acid Hybridization To Combed DNA Fibers

In the following part we describe the development of a
CGH-test providing a resolution at the single gene level. This
test can be fully automated and broadly used in clinical
settings.

The new test is based on the idea that, instead of entire
chromosomes, specific target nucleic acids (DNAs or RNAs) are
immobilized on a supportive matrix, such as glass or plastic

materials, in any desirable geometric format. The number of
target nucleic acid (TNA-) spots and the sequence complexity of
each TNA-spot can be chosen with regard to the specific goals
of a test (see the application examples below). The number of
TNA-spots included in a given matrix-CGH test may vary from a
few spots to hundreds or even thousands of spots (for potential
applications see below).

A typical TNA-spot may contain DNA from a single cosmid
representing a gene or part of a gene of interest or it may
contain a complex mixture of DNA representing a chromosome
segment or even an entire chromosome of interest. In the
latter case a matrix CGH test would not provide a resolution
superior to the resolution of CGH to reference metaphase
chromosomes. In the following we will construct our
considerations mainly to the development of a matrix CGH test
with the highest conceivable resolution, i.e., a test to detect
copy number changes in a set of selected genes.

A matrix with TNA-spots as described above can be used to
test tumor or other test DNAs for genetic imbalances down to
the kbp-range. For this purpose, the hybridization probe
consisting of a 1:1 mixture of differently labeled test and
reference genomic DNAs (or RNA- or cDNA-preparations) is
hybridized under suppression conditions against the set of
immobilized TNA-spots. Measurements of the fluorescence ratio
on each individual TNA-spot should provide an estimate of the
copy number representation of the respective target sequences
in the test DNA (or test RNA) as compared to the reference DNA
(or reference RNA) (for further details of measurements see
below).

The successful development of such a test depends on three
requirements, namely the ability to firmly immobilize target
nucleic acids on a supportive matrix, e.g., glass or plastic, a
low autofluorescence of the matrix, and a sufficiently high
signal/noise ratio for sequences specifically hybridized to a
given TNA-spot. Notably, in CGH experiments with two

differently labeled genomic DNAs the fraction of labeled DNA
fragments, which are specific for a given TNA-spot, is
generally very small. Any non-specific attachment of labeled
sequences or detection reagents to the matrix may impair or
even inhibit the measurement of meaningful fluorescence ratios.
Such adverse effects may become a limiting factor in attempts
to measure fluorescence ratios on entire TNA-spots,
particularly in cases where the specific signal is relatively
small as compared to background.

In order to minimize background problems in fluorescence
ratio measurements we make use of a procedure called "molecular
combing" (Bensimon et al. 1994), which is relied upon and
incorporated in its entirety by reference herein. By this
procedure DNA target fibers can be extended and aligned in
parallel like hairs by the use of a comb. To this end, the DNA
fibers are attached at one end to a solid surface, "combed" by
a receding air-water interface, and finally immobilized on the
drying surface. (Bensimon et al. 1994). In this way the DNA
of each TNA-spot can be represented by a series of "combed"
target DNA fibers, all representing a specific DNA sequence of
interest, e.g., a cosmid clone from a gene of interest. Using
standard fluorescence in situ suppression hybridization
techniques under appropriate stringency conditions,
complementary sequences present in the hybridization probe
hybridize specifically to these target sequences.

For each TNA-spot, fluorescence is separately recorded for
both fluorochromes on a series of individual combed target DNA
fibers using an appropriate camera, such as a CCD-camera. For
evaluation, each target sequence is enclosed in a narrowly
adapted rectangular field to determine the fluorescence of the
fluorochromes applied in the labeling/visualization of the
hybridized DNA fragments. Background fluorescence is measured
in the immediate neighborhood of each measured target sequence
and subtracted. After background substraction, the
fluorescence ratio is determined for each individual target

sequence with a microprocessor. Foe each TNA-spot the
variation of fluorescence ratios obtained for a number of
combed target DNA fibers is determined. Target sequences,
which are represented in normal copy numbers in both the test
5   and the reference gnomic DNA, are used to obtain reliable
thresholds for ratios indicative for increased or decreased
copy number representation. For TNA-spots representing target
sequences, which are over- or underrepresented in the test-DNA,
one should then obtain correspondingly increased or decreased
10 fluorescence ratios, while balanced regions should yield a
ratio within the limits of control experiments. Since
fluorescence ratios are recorded from individual target DNA
fibers present in the TNA-spots, the new test is inert against
variations in the total amount of target-DNA in each TNA-spot.

15   In model experiments, probe mixtures consisting of
different ratios of biotin labeled and digoxigenin labeled
cosmid sequences (e.g., 1:1 (20 ng + 20 ng), 2:1 (20 ng + 10
ng), 5:1 (20 ng + 4 ng) and 10:1 (20 ng + 2 ng)) were prepared.
For each chosen ratio of sequences in the probe mixture, the
20 same cosmid was used as target sequence. Following comparative
hybridization some fifty target DNA fibers were evaluated as
described in the Examples. The results demonstrate highly
significant differences of mean fluorescence ratio values
obtained for the different probe mixtures.

25   The use of individual target DNA fibers provides the
possibility of another approach for the evaluation of a CGH
experiment. Instead of measuring fluorescence intensities from
entire target DNA fibers, it is possible to simply count
individual fluorescence hybridization dots. Each such dot
30 presumably represents the hybridization of an individual probe
DNA fragment of a few hundred bp. The coverage of the combed
target DNA sequence with differently colored dots from the 1:1
hybridization mixture of genomic DNAs is a stochastic event.
The dot numbers for the two differently labeled genomic DNAs
35 counted over a series of target DNA DNA fibers therefore

reflect the copy number representations of the target DNA sequence in the two genomic DNAs used for comparative hybridization.

Notably, using such an approach it is not required to achieve sufficient signal from the different labeled genomic DNAs over each target sequence to allow meaningful fluorescence measurements. Consider that the test genomic DNAs on average yields five specific hybridization dots, while the reference genomic DNA yields one specific dot per target DNA sequence. Then the conclusion seems valid that the copy number representation of the target sequence in question is five times higher in the test genomic DNA.

The ratio of dots from test and reference genomic DNA counted over a series of target DNA DNA fibers contained in a given TNA-spot may deviate from the actual ratio of copy number representation of the target sequence in the test and reference genomic DNAs for a number of reasons. The differently labeled DNAs in the 1:1 hybridization mixture should be digested to the same size distribution. The number of background dots in the vicinity of the target sequence should be approximately the same for both the test and reference genomic DNA. If necessary, the number of dots, which are expected to result from chance background dots on the target DNA sequence, should be calculated from the area surrounding the target DNA sequence and subtracted from the overall number of counted dots. Only those background dots that are ordered exactly in line along a combed target DNA sequence can be confused with actual signal dots. The proposed approach to measure dots along combed DNA target DNA fibers thus helps strongly to minimize the number of background dots, which could deteriorate the accuracy of counting the number of specific hybridization dots along a series of combed target DNA fibers. This is a decisive advantage of the proposed test as compared to a test where fluorescence ratios are determined from an entire DNA-spot built up by a large number of non-combed DNA fibers.

Further, it should be emphasized that the proposed test allows the comparison of fluorescence or dot number ratios over a number of TNA-spots. TNA-spots, which contain combed Target DNA fibers present in equal copy number in both the test and
5   reference genomic DNA, can serve to standardize the fluorescence ratios and dot ratios, respectively, not only with regard to the internal standardization in each individual TNA-spot given by the fluorescence or dot number obtained from the reference genomic DNA, but also with regard to the
10 standardization of the data between different TNA-spots representing combed target DNA fibers present in equal and different copy numbers in the test and the reference genomic DNA.

In case that each target DNA DNA fiber is covered with
15 numerous specific dots, the target fibers can be easily distinguished as a linear array of dots. The unequivocal identification of the combed target DNA fibers is an absolutely essential requirement to count small numbers of dots, where a dot number ratio (or difference obtained by subtraction) is
20 only meaningful when obtained from a series of target DNA fibers. Therefore the target fibers need to be visualized by other means when the number of dots is small, including specific DNA fluorochromes with an emission spectrum distinguishable from the fluorochromes used for the
25 identification of hybridized DNA (or RNA) fragments.

Target DNA fibers can also be visualized by the admixture of labeled target DNA to the hybridization mixture. In this case a third label is required in addition to the two labels for the test and reference genomic DNA. The admixture of
30 labeled target sequences has to be carefully adjusted in order to avoid too much suppression of the hybridization of the labeled target sequences present in the test and reference genomic DNAs. This problem can be avoided, if linkers are adapted to the target DNA sequence, which can be visualized by
35 hybridization with a specific linker probe. The combed target

DNA sequence in question would then be embraced by two
fluorescently labeled linker sequences. Alternatively, target
DNA sequences can be cloned in the presence of hapten modified
nucleotides, such as BrdU, and visualized immunocytochemically.

5   For the choice of useful target DNA sequences one should
take into account that sequences that contain interspersed
repetitive signals, e.g., Alu elements, require suppression
hybridization, e.g., with an excess of unlabeled Cot1-DNA. It
might be preferable to use target DNA fibers, which are
10   entirely specific for the genomic region in question, or to
construct target DNA fibers devoid of interspersed repetitive
elements.

          CGH on combed DNA fibers bears the potential for an
ultrahigh resolution CGH. Consider the following scenario: A
15   DNA fiber with known DNA sequence contains a target region of
interest comprising a few hundred base pairs. We assume that
the copy number of this target region is variable and may be
higher (or lower) in the test genomic DNA as compared to the
reference genomic DNA. We assume further that the positions of
20   the target and control regions along the DNA fiber are
precisely mapped and that the DNA fiber is engineered in a way
that its 5' - 3' orientation can be visualized, e.g., by probes
to linker adapters of different size. In such a scenario the
target region could be mapped by fractional length measurements
25   on each DNA fiber. Accordingly, one could identify dots which
represent hybridization events to the target region and count
the number of such events on a series of DNA fibers. This
number should correlate with the representation of target
sequences in the genomic test and reference DNA. In case that
30   the number target sequences is increased in the test genomic
DNA as compared to the reference genomic DNA, the respective
dot ratio should be increased over this region in contrast to
other regions of the DNA fiber, for which we assume an equal
copy number representation in both the test and reference
35   genomic DNA.

As a model system one could prepare PCR-amplified probes
for two regions of a given DNA-fiber. DNA aliquots from each
probe could be differently labeled and various hybridization
mixtures prepared, where the differently labeled aliquots are
present in different ratios. By counting the hybridization
dots over the two target regions on a series of DNA fibers, one
could determine to which extent the dot number ratios reflect
the ratios of the differently labeled probe aliquots. This
example may illustrate that an approach based on dot counting
may be feasible to determine the copy number representation of
very small target sequences where fluorescence ratios are no
longer valid. This advantage is reemphasized by the following
scenario.

Consider single-stranded DNA sequences containing a small
(300 bp) target sequence of interest. Four CGH probe mixtures
with differently labeled complementary 300 bp sequences are
applied. On each individual fiber the target sequence of
interest can only hybridized once, i.e., it will bear a dot of
one color only. If the target is represented by double-
stranded, denatured DNA, both strands serve as targets for a
denatured double-stranded probe. Accordingly, the target
region can hybridize with two 300 bp fragments at most. The
resulting dots are either of the same color or of different
color. While the measurement of a fluorescence ratio over the
target region of a single fiber is obviously meaningless, the
frequency with which the target region is covered by dots of
the same (e.g., green or red) or of different colors (e.g.,
yellow) in a series of DNA fibers should be highly informative
with regard to the frequency of differently labeled target
sequences in the hybridization mixture.

Alternatively, DNA representing a target sequence of a few
hundred base pairs could be fixed to the matrix the target spot
and a fluorescence ratio could be determined from the entire
spot. In this case, however, background could become a major
problem. From these considerations we conclude that a dot

counting approach performed on a sufficient number of individual DNA fibers may be superior or even the only feasible way in case of ultrahigh resolution CGH.

In contrast to the largely variable size, form, and relative position of individual target chromosomes considered in CGH on reference metaphase spreads, the length and orientation of a combed target DNA fibers can be strictly controlled. These defined patterns of combed target DNA fibers strongly facilitate their fully automated evaluation. In case of a large number of TNA-spots, a colored print-out of fluorescence ratio or dot ratio measurements is recommended to facilitate the investigator's recognition of genomic regions, which are over- or under represented. In this print-out, each TNA-spot is represented by a colored spot. One color should reflect TNA-spots with the range of fluorescence ratios apparently representing sequences present in balanced copy number in the test-DNA, a second color should reflect TNA-spots with sequences present in increased copy number, while a third color should reflect TNA-spots with sequences present in decreased copy number. If desirable, color intensity may reflect the relative extent of over- or underrepresentation.

For a series of TNA-spots containing physically mapped sequences, the color spots could be arranged in a way that reflects mapping positions on chromosomes. For example, a linear array of colored spots could represent the order of clones in a contig used for combed DNA fiber CGH. In case that the whole chromosome complement is represented by TNA-spots, the resulting color spots could be ordered as 24 linear arrays (representing chromosomes 1-22, X and Y). The color spots within a given array then could represent the physical order of clones within the respective chromosome. The investigator then is enabled to see at one glance which chromosomes or chromosomal subregions are present in balanced, increased or decreased copy number.

Potential Applications Of CGH On A

Matrix With Combed Target DNA Fibers

In this section we will consider potential applications of CGH on combed target DNA fibers. In addition to potential studies of tumor DNA samples, we will also include applications in clinical genetics and cytogenetics.

Tests with gene specific TNA-spots would allow the rapid screening of test DNAs from tumor samples for copy number changes of specific genes. Given the appropriate equipment for the automated evaluation of TNA-spots, it may become feasible to evaluate DNA spot matrices with numerous spots representing the copy number representation of an entire set of genes of interest at a reasonable price.

A survey of whole genomes at the highest possible level of resolution would require such a high number of spots that such an approach appears impractical. For practical purposes, it may be advantageous to perform a survey of a test genomic DNA with unknown gains and losses by a series of tests with increasing resolution, starting by CGH on metaphase spreads and subsequently homing in on specific chromosome segments. In addition to CGH on reference chromosomes, matrices with spots representing composite DNA sequences of entire chromosomes, chromosome arms or bands may be applied. CGH performed directly on target DNA fibers provides the ultimate level of resolution for CGH and is only reasonable in cases where the screening of a very specific subset of target DNA sequences for copy number changes in a test genomic DNA is required. The following examples may illustrate how such a strategy could be applied.

CGH on reference chromosome spreads may reveal, for example, the non-random loss of a certain chromosome segment for a certain tumor entity. A consensus region can be defined by the comparison of all tumors showing this deletion. However, a considerable fraction of the tumors may not show any detectable deletion at this level of resolution (>10 Mbp). To

screen these tumors for much smaller deletions, a matrix with
TNA-spots representing the consensus region could be used. The
resolution would depend on the size and linear, genomic
distance of the target sequences represented by a given matrix.
In many cases it should be sufficient for screening purposes to
represent a chromosome region by a series of TNA-spots, where
each spot defines for example a cosmid sequence a few hundred
kb apart from the target DNA sequence contained in the next
TNA-spot.

To achieve the highest, possible resolution even a whole
contig can be represented by a series of TNA-spots.
Fluorescence ratio measurements performed on a high resolution
matrix representing a region of interest should help to define
the cosmids, which represent the smallest deletion detectable
in this region of interest for a whole series of genic test
DNAs obtained from patients with a specific tumor entity. This
minimum deletion could be confirmed by FISH of the respective
cosmids to tumor nuclei.

In this way, efforts of positional cloning of a suspected
tumor suppressor gene could be strongly facilitated. For such
a purpose it would not be necessary to know the precise linear
order of the cosmid clones representing the chromosome segment
of interest. If, say, three TNA-spots with decreased
fluorescence ratio would define the minimal detectable
deletion, one would expect that FISH with these three clones to
extended chromatin fibers would confirm their vicinity. The
latter approach could also be used to map the linear order of
these cosmids.

Similarly, matrices with physically mapped cosmids
representing a chromosomal subregion could help to define
amplified regions. Positional cloning of the genes involved in
amplifications would be greatly facilitated, if the extension
of such amplifications could be precisely mapped. Consider
that a chromosome band has been identified as the source of the
amplified sequences by CGH to reference metaphase spreads.

Applying a matrix with a series of physically mapped clones
representing the chromosome band in question should yield
increased fluorescence ratios for any TNA-spot representing DNA
sequences, which are in fact amplified. Normal fluorescence
5   ratios should be measured for all other TNA-spots. The extent
of the amplification could then be defined by the two TNA-spots
with increased fluorescence ratios, which represent the most
proximally and most distally mapping clones.

In the future, matrices containing TNA-spots with combed
10   target DNA fibers for the copy number representation of
oncogenes and tumor suppressor genes can be developed. The
choice of the target sequences for a given matrix will depend
on the tumor entity and the demands of the test. For example,
matrices can be specifically developed to identify gains or
15   losses with prognostic value (e.g., N-mye amplifications or
1p36- deletions in neuroblastomas). For some tumors, e.g.,
colorectal tumors, knowledge about the relevant genes involved
in tumor initiation and progression seems already sufficiently
advanced to consider the development of such a strategy. For
20   other tumors we still lack such knowledge. CGH on combed DNA
fibers may help to obtain such knowledge in the future and to
perform large scale tests performed with the aim to correlate
the patterns of relative copy number changes of genomic DNA
sequences (and changes in the number of specific mRNAs) in
25   tumor cells with the clinical course of the disease. Ideally,
matrices should be developed, which contain TNA-spots
representing all genes that are relevant for the biological
properties of the tumor entity in question.

High resolution matrices could also open new avenues in
30   clinical cytogenetics. Two examples may be sufficient to
demonstrate the range of possible applications. A CGH test-
matrix could be developed to screen DNA from patients with
phenotypes suspicious for unbalanced chromosome aberrations.
Taking into account that unbalanced rearrangements often
35   include terminal chromosome segments, a CGH-matrix containing

TNA-spots with combed target DNA fibers representing cloned
sequences from each individual chromosome end may become a
great clinical value. As another example, consider the case of
a carrier-analysis for X-linked recessive diseases. A boy, who
suffers from Duchenne muscular dystrophy, may be the first
victim of that disease in a family. In some 60% of the cases a
deletion can be found as the cause of the mutation. The
question has then to be answered whether his disease is due to
a germ cell mutation or whether his mother is already a
carrier. The consequences for genetic counseling are totally
different and in the latter case other female members, e.g.,
the sisters of the boy's mother, may also be concerned about
their carrier-status. A matrix with a series of cosmids
spanning the entire dystrophin gene could potentially provide a
reliable and automated procedure for carrier screening in
deletion prone cases of DMD.

Comparative RNA Hybridization On Combed

DNA Fibers: A New Method For The Assessment

Of Expression Levels Of Tumor Relevant Genes

While CGH on combed DNA fibers should provide information
on deleted or amplified genes, it would not detect the
silencing or overexpression of genes in tumor cells as compared
to their normal counterparts. We propose an approach to study
the expression status of genes by comparative RNA hybridization
on combed DNA fibers. Consider for example a scenario where an
amplification is detected in a given tumor entity. Amplicons
may be large and contain several genes. It may not be clear
which gene(s) of these genes are strongly expressed. A DNA-
spot matrix containing combed cDNA-fibers for the coding
sequences of all genes in question can be used for comparative
nucleic acid hybridization with differently labeled RNA-
preparations (or corresponding cDNA preparations) from the
tumor and a normal reference tissue. The resulting
fluorescence or dot number ratios then can provide insight in
the (relative) expression status of the tested genes in the

tumor as compared to normal tissue. The same approach could be used to detect point mutations which interfere with the description of a gene.

This invention will now be described in greater detail in the following Examples.

EXAMPLE 1

Preparation Of Glass Slides With Combed DNA-fibers

Any cloned DNA, PCR-amplified DNA or other purified DNA can be used for DNA-combing depending on the purpose of the comparative hybridization experiment. The amount of DNA necessary per TNA-spot is small, since only a small number of DNA-fibers is needed for evaluation is combed fiber CGH experiments (for example, 1 ng of cosmid DNA (40kb) contains 2.4×102 molecules). In the following we describe a typical experiment with cosmid DNA.

Attachment of target DNA fibers requires glass surfaces pretreated by salinization as described previously (Bensimon et al. 1994). Prior to their attachment, cosmids can be stained for 1 hour at room temperature with YOYO-1 (Molecular probes; Cat. No. Y-3601). Staining solution was freshly prepared as follows: 1^ml DNA (1mg/ml) + 33^ml YOYO-1 diluted 1:100 in $T_{40}E^{\wedge}_2$ + 66 ^ml $T_{40}E^{\wedge}_2$; $t_{40}E^{\wedge}_2$ buffer contains 40 mM Tris-Acetate, 2 mM EDTA, pH 8.0). YoYo-1 stained cosmid DNA was diluted in 50 mM MES-buffer, pH 5.5, and then attached to the glass surface (note that the pH is a most critical point for successful attachment).

In our present experiments, each glass slide (22 x 22 mm) contained only one type of combed DNA fibers. Where appropriate, several glass slides containing DNA fibers with different sequences were processed in parallel.

For the preparation of a single matrix with a series of TNA spots with combed DNA fibers, salinization of the glass surface can be restricted to the areas selected for the positions of the TNA spots. A series of droplets, each containing DNA fibers with the required DNA target sequence for

a given area, can be put on these preselected areas. Target
fibers contained in each droplet are allowed to attach to the
surface. The excess fluid with non-attached fibers is removed
taking care that the surface is kept wet.

Following DNA fiber attachment, combing was carried out as
described (Bensimon et al. 1994). Microscopic visualization of
YoYo-1 stained fibers allowed a control of the density and
direction of attached fibers. Slides with combed DNA fibers
were baked at 60° C for at least 4 hours and treated with a
"blocking solution" consisting of 3% BSA in 2xSSC at 37°C for
30 min. After a brief wash with 2xSSC, slides were put through
a series of 70%, 90%, 100% E:OH, 2 min each and air dried.
Storage of the dried slides is recommended in sealed boxes at
+4°C.

EXAMPLE 2

Comparative Hybridization To Combed DNA-fibers

Comparative genomic hybridization (CGH) to combed DNA
fibers was essentially carried out as described elsewhere for
CGH to metaphase chromosomes (du Manoir et al. 1993, 1995).
Briefly, test and reference genomic DNAs were nicktranslated
with biotin and digoxigenin, respectively. Alternatively, the
DNAs can be labeled directly with appropriate fluorochrome
conjugated nucleotides. Combed DNA fibers were denatured for 2
min. at 72°C in 70% FA/0.6xSSC, pH7.0. Thereafter, slides were
put through a series of ice cold ErOH (70%, 90%, 100%) and air
dried. Ten ^ml hybridization mixture (containing 500 ng each
of the test and reference genomic DNA, 50 ^mg of Cot1 fraction
of human DNA (BRL/Life Technologies) and 55 ^mg sonicated
salmon testes DNA (Sigma) in 50% formamide, 1xSSC and 10%
dextrane sulphate) were put on each glass slide (22 x 22 mm)
with combed DNA fibers. Another slightly smaller glass slide
(18 x 18 mm) was put on top and sealed with fixogum.

Hybridization was carried out overnight at 37°C. Washing
and detection procedures for biotin and digoxigenin labeled
sequences were carried out as described (Lichter and Cremer

1992, du Manoir et al. 1993, 1995) with minor modifications.
Slides were washed 3 x 5 min. with 50% FA/SSC and another 3 x 5
min. with 2SSC at room temperature. Following equilibration in
4xSSC/0.1% Tween 20 at 37°C slides were incubated for 30 min.

5   with 3% BSA/4SSC/0.1% Tween 20 at 37°C (a blocking step to
reduce background). Slides were then washed for 5 min. in
4xSSC/0.1% Tween 20 at 37°C and incubated with avidin DCS
conjugated to FITC (Vector Laboratories) for 45 min. at 37°C to
visualize biotin labeled probes.

10 Digoxigenin-labeled probes were detected by incubation
with mouse-anti-digoxin IgG antibodies as the primary antibody
(Sigma), followed by incubation with a sheep-anti-mouse IgG
antibody conjugated with the fluorochome CY3. In between these
steps slides were washed 5 min. each in 4xSSC/0.1% Tween 20 at

15 37°C. If necessary, avidine-FITC signals were amplified as
described (Pinkel et al. 1986). Finally, slides were air-dried
and mounted in Vectashield (Vector Laboratories) as an
antifade.

In a series of model experiments, probe mixtures
20 consisting of different ratios of biotin labeled and
digoxigenin labeled cosmid sequences (e.g., 1:1 (20 ng + 20
ng), 2:1 (20 ng + 10 ng), 5:1 (20 ng + 4 ng) and 10:1 (20 ng +
2 ng) were used for in situ hybridization to combed DNA fibers
representing the same cosmid as target sequence.

25 **EXAMPLE 3**

Evaluation Of Combed DNA Fibers

Subjected To Comparative Hybridization

(A) Image acquisition

Grey-level images were recorded with a cooled black and
30 white CCD camera (Photometrics) separately for each
fluorochrome. Optimal exposure times and optical settings were
established empirically and then kept constant for the entire
set of DNA fibers recorded for a given experiment. Images were
stored under FITS-format.

(B)   Image processing

   (1)   Measurement of fluorescence ratios

   Digital images were processed by NIH image (version 1.59b9). A rectangular mask was adapted to each DNA fiber. Within the mask two integrated fluorescence values were obtained as the sum of the grey level values for all pixels. Background fluorescence intensity was determined after shifting the mask to the immediate neighborhood of a given DNA fiber. Fluorescence intensities were corrected by background subtraction. The fluorescence ratio for each fiber was calculated by dividing the corrected FITC fluorescence intensity with the corrected Cy3 fluorescence intensity. Fifty DNA fibers were evaluated in a typical experiment to calculate the mean fluorescence ratio value.

   (2)   Dot counting

   Notably, the labeling observed along combed DNA fibers in the experiments described above is not homogeneous. Instead signal dots probably representing hybridized labeled DNA fragments can be distinguished on these fibers. For dot counting, digital images were thresholded and gravity centers were determined.

   In summary, this invention provides procedure termed combed
DNA fiber comparative hybridization (CFCH). For this new procedure, there is provided

     1)   A possibility to measure the relative copy
      number representation of DNA sequences from
      labeled genomic test DNA (e.g. tumor DNA, DNA
      from patients with imbalanced types of
      chromosome aberrations) as compared to the copy
      number representation in differently labeled
      normal genomic DNA with a resolution in the
      kilobase pair range, i.e. at the level of single
      genes. In contrast, CGH as described before can

detect copy number changes only in the megabase pair range.

2)   A possibility to measure copy number differences in mRNAs transcribed from specific genes by comparative hybridization of a mixture of mRNAs prepared from test and reference cells.

3)   The arrangement of target DNA spots with combed DNA fibers representing genomic or cDNA test sequences of interest on a suitable matrix for the simultaneous testing of multiple sequences for their relative copy number representation in the hybridization mixture in a geometrical format, which makes such an arrangement particularly useful for automated evaluation.

The invention is particularly useful for the detection of genetic diseases in eucaryotic cells.

REFERENCES

Bensimon, A., Simon, A., Chiffaudel, A., Croquette, V., Heslot, F., Bensimon, D. Science 265:2096-2098 (1994).

du Manoir, S., Speicher, M.R., Joos, S., Schrock, E., Popp, S., Dohner, H., Kovacs, G., Robert-Nicoud, M., Lichter, P., Cremer, T. Hum. Genet. 90:590-610, (1993).

du Manoir, S., Schrock, E., Bentz, M., Speicher, M.R., Joos, S., Ried, T., Lichter, P. Cremer, T. Cytometry 19:27-41 (1995).

Joos, S., Scherthan, H., Speicher, M.R., Schlegel, J., Cremer, T., Lichter, P. Hum. Genet 90:584-589 (1993).

Isola, J., DeVries, S. Chu, L., Ghazvini, S., Waldman, F. Am. J. Path. 145:1301-1308 (1994)

Kallioniemi, A., Kallioniemi, O.P., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F., Pinkel, D. Science 258:818-821 (1992).

Kallioniemi, O.P., Kallioniemi, A., Piper, J., Isola, J., Waldman, F.M., Gray J.W., Pinkel, D. Genes, Chromosomes and Cancer 10:231-243 (1994).

Kallioniemi, O.P., Kallioniemi, A., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F., Pinkel, D. Sem. Cancer Biol. 4:41-46 (1993).

Kallioniemi, A., Kallioniemi, O.P., Piper, J., Tanner, M., Stokke, T., Chen. L., Smith, H.S., Sudar, D., Pinkel, D., Gray, J.W., Waldman, F. Proc. Natl. Acad. Sci. USA 91:2156-2160 (1994).

Lichter, P., Cremer, T., Borden, J., Manuelidis, L., Ward, D.C. Hum. Genet. 80:224-234 (1988).

Lichter, P., Cremer, T.: Chromosome analysis by non-isotopic in situ hybridization. In: Human cytogenetics: A practical approach; eds.: Rooney, D.E., Czepulkowski, B.H., IRL Press, Oxford:157-192 (1992).

Muleris, M. Almeida, A., Dutrillaux, A.M., Pruchon, E., Vega, F., Delattre, J.Y., Poisson, M. Malfoy, B., Dutrillaux, B. Genes, Oncogene 9:2717-2722 (1994).

Pinkel, D., Landegent, J.; Collins, C. Fuscoe, J., Segraves, R. Lucas, J., Gray, J. Proc. Natl. Acad. Sci. U.S.A. 85:9138-9142 (1988).

Piper, J., Rutovitz, D., Sudar, D., Kallioniemi, A.,
5 Kallioniemi, O.P., Waldman, F.M., Gray, J.W. Pinkel, D. Cytometry 19:10-26 (1995).

Ried, T., Petersen, L., Holtgreve-Grez, H., Speicher, M.R., Schrock, E., du Manoir, S., Cremer, T. Cancer Res. 54:1801-1806 (1994).

10 Schrock, E., Thiel, G., Lozanova, T., du Manoir, S., Meffert, M.C., Jauch, A., Speicher, M.R. Nurnberg, P. Vogel, S., Janisch, W., Donis-Keller, H., Ried, T., Witkowski, R., Cremer, T. Am J. Path. 144:1203-1218 (1994).

Speicher, M.R., du Manoir, S., Schrock, E., Holtgreve-
15 Grez, H., Schoell, B., Lengauer, C., Cremer, T., Ried, T. Hum. Mol. Genet. 2:1907-1914 (1993).

Speicher, M.R., Prescher, G., du Manoir, S., Jauch, A., Horsthemke, B., Bornfeld, N., Becher, R., Cremer, T. Cancer Res. 54:3817-3823 (1994).

20 Speicher, M.R., Jauch, A., Walt, H., du Manoir, S. Ried, T., Jochum, W., Sulser, T. Cremer, T. Am J. Path. 146: 1332-1340 (1995).

What is claimed is :

1. A method for detecting and quantifying the presence of a nucleic acids sequence in a genome comprising:

(A) immobilizing by combing the target nucleic acids sequence on a supportive matrix,

(B) contacting the combed nucleic acids sequence with a 1:1 mixture of differently labeled nucleic acids probe to be tested and reference nucleic acids probe to form hybridization,

(C) counting individual labeled hybridization dots located on combed nucleic acids sequence,

(D) determining the number of the nucleic acids sequence to be detected and quantified by the calculation of the ratio between the total dot number from the test probe and the total dot number from the reference probe.

2. A method according to claim 1, wherein the combing of the target sequence comprises:

(A) anchoring the end of said sequence to a supportive matrix,

(B) contacting the anchored sequence with a liquid to form a meniscus,

(D) combing said sequence by a receding air-water interface.

3. A method according to claim 2, wherein said liquid has a pH about 5 to about 10 to form a meniscus.

4. A method according to claims 1 to 3, wherein the target sequence is labeled.

5. A method according to claims 1 to 4, wherein the combed target sequence is DNA and test and reference probe are RNA- cDNA- or genomic DNA-preparations.

6. A method according to anyone of claims 1 to 5, wherein the nucleic acids sequence to be detected and quantified is a human oncogene.

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6   C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6   C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 96 17958 A (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 13 June 1996 see the whole document ---	1-6
X	WO 93 18186 A (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 16 September 1993 see the whole document ---	1,4-6
X	WO 95 21939 A (INSTITUT PASTEUR) 17 August 1995 see the whole document, especially page 14 line 12 - page 17, line 6 and pages 19-21 ---	1-3
	-/--	

X	Further documents are listed in the continuation of box C.		X	Patent family members are listed in annex.

* Special categories of cited documents :

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document but published on or after the international filing date

'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

'O' document referring to an oral disclosure, use, exhibition or other means

'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

'&' document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 February 1997	04.03.97

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	De Kok, A

Form PCT ISA 210 (second sheet) (July 1992)

1

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CELLULAR AND MOLECULAR BIOLOGY, vol. 41, no. 7, November 1995, PARIS FR, pages 925-931, XP000197050 C. DESMAZE ET AL.: "In situ hybridization of fluorescent probes on chromosomes, nuclei or stretched DNA: applications in physical mapping and characterization of genomic rearrangements" see the whole document ---	1-6
X	SCIENCE, vol. 270, no. 5235, 20 October 1995, LANCASTER, PA US, pages 467-470, XP000644675 M. SCHENA ET AL.: "Quantitative monitoring of gene expression patterns with a complementary DNA microarray" see the whole document ---	1-6
A	WO 95 09929 A (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 13 April 1995 see the whole document -----	1

1

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9617958 A	13-06-96	NONE		
WO 9318186 A	16-09-93	CA	2131543 A	16-09-93
		EP	0631635 A	04-01-95
		JP	7505053 T	08-06-95
WO 9521939 A	17-08-95	FR	2716206 A	18-08-95
		FR	2716263 A	18-08-95
		AU	1712095 A	29-08-95
		AU	1814595 A	29-08-95
		EP	0744028 A	27-11-96
		EP	0743988 A	27-11-96
		WO	9522056 A	17-08-95
WO 9509929 A	13-04-95	US	5472842 A	05-12-95
		CA	2171110 A	13-04-95
		EP	0722507 A	24-07-96